DISTRIBUTED STOCHASTIC SUBGRADIENT PROJECTION ALGORITHMS FOR CONVEX OPTIMIZATION

S. SUNDHAR RAM, A. NEDIĆ, AND V. V. VEERAVALLI *

Abstract. We consider a distributed multi-agent network system where the goal is to minimize a sum of convex objective functions of the agents subject to a common convex constraint set. Each agent maintains an iterate sequence and communicates the iterates to its neighbors. Then, each agent combines weighted averages of the received iterates with its own iterate, and adjusts the iterate by using subgradient information (known with stochastic errors) of its own function and by projecting onto the constraint set.

The goal of this paper is to explore the effects of stochastic subgradient errors on the convergence of the algorithm. We first consider the behavior of the algorithm in mean, and then the convergence with probability 1 and in mean square. We consider general stochastic errors that have uniformly bounded second moments and obtain bounds on the limiting performance of the algorithm in mean for diminishing and non-diminishing stepsizes. When the means of the errors diminish, we prove that there is mean consensus between the agents and mean convergence to the optimum function value for diminishing stepsizes. When the mean errors diminish sufficiently fast, we strengthen the results to consensus and convergence of the iterates to an optimal solution with probability 1 and in mean square.

Key words. Distributed algorithm, convex optimization, subgradient methods, stochastic approximation.

AMS subject classifications. 90C15

1. Introduction. A number of problems that arise in the context of wired and wireless networks can be posed as the minimization of a sum of functions, when each component function is available only to a specific agent [23, 25, 26]. Often, it is not efficient, or not possible, for the network agents to share their objective functions with each other or with a central coordinator. In such scenarios, distributed algorithms that only require the agents to locally exchange limited and high level information are preferable. For example, in a large wireless network, energy is a scarce resource and it might not be efficient for a central coordinator to learn the individual objective functions from each and every agent [23]. In a network of databases from which information is to be mined, privacy considerations may not allow the sharing of the objective functions [34]. In a distributed network on a single chip, for the chip to be fault tolerant, it is desirable to perform the processing in a distributed manner to account for the statistical process variations [32].

We consider constrained minimization of a sum of convex functions, where each component function is known partially (with stochastic errors) to a specific network agent. The algorithm proposed builds on the distributed algorithm proposed in [19] for the unconstrained minimization problem. Each agent maintains an iterate sequence and communicates the iterates to its neighbors. Then, each agent averages the received iterates with its own iterate, and adjusts the iterate by using subgradient information (known with stochastic errors) of its own function and by projecting onto the constraint set. The inter-agent information exchange model is a synchronous and delayless version of the computational model proposed by Tsitsiklis [30]. The

algorithm is *distributed* since there is no central coordinator. The algorithm is *local* since each agent uses only locally available information (its objective function) and communicates locally with its immediate neighbors.

Related to this work are the distributed incremental algorithms, where the network agents sequentially update an iterate sequence in a cyclic or a random order [5, 12, 16, 23, 25]. The effects of stochastic errors on these algorithms have been investigated in [3, 9, 14, 17, 23, 25, 28]. In an incremental algorithm, there is a single iterate sequence and only one agent updates the iterate at a given time. Thus, while being distributed and local, incremental algorithms differ fundamentally from the algorithm studied in this paper (where all agents update simultaneously). Also related are the optimization algorithms in [2, 31]. However, these algorithms are not local as the complete objective function information is available to each and every agent, with the aim of distributing the processing.

The work in this paper is also related at a much broader level to the distributed consensus algorithms [2,11,13,15,18,20,21,29–31,33]. In these algorithms, each agent starts with a different value and through local information exchange, the agents eventually agree on a common value. The effect of random errors on consensus algorithms have been investigated in [10, 13, 15, 33]. In addition, since we are interested in the effect of stochastic errors, our paper is also related to the literature on stochastic subgradient methods [6–8].

We consider general stochastic errors that have uniformly bounded second moments and obtain bounds on the limiting performance of the algorithm in mean for diminishing and non-diminishing stepsizes. When the means of the errors diminish, we prove that there is mean consensus between the agents and mean convergence to the optimum function value for diminishing stepsizes. When the mean errors diminish sufficiently fast, we strengthen the results to consensus and convergence of the iterates to an optimal solution with probability 1 and in mean square.

Our work expands the multi-agent distributed optimization framework studied in [19]. The new contributions are: 1) the study of the effects of stochastic errors in subgradient evaluations; 2) the consideration of constrained optimization problem within the distributed multi-agent setting. The presence of the constraint set complicates the analysis as it introduces non-linearities in the system dynamics. The non-linearity issues that we face have some similarities to those in the constrained consensus problem investigated in [20], though the problems are fundamentally different. The presence of subgradient stochastic errors adds another layer of complexity to the analysis as the errors made by each agent propagate through the network to every other agent and also across time, making the iterates statistically dependent across time and agents.

The rest of the paper is organized as follows. In Section 2, we formulate the problem, describe the algorithm and state our basic assumptions. In Section 3 we state some results from literature that we use in the analysis, while in Section 4, we derive two important lemmas that form the backbone of the analysis. In Section 5, we study the convergence properties of the method in mean, and in Section 6 we focus on the convergence properties with probability 1 and in mean square. Finally, we discuss some implications and provide some concluding remarks in Sections 7 and 8.

2. Problem, algorithm and assumptions. In this section, we formulate the problem of interest and describe the algorithm that we propose. We also state and discuss our assumptions on the agent connectivity and information exchange.

2.1. Problem. We consider a network of m agents that are indexed by $1, \ldots, m$. Often, when convenient, we index the agents by using set $V = \{1, \ldots, m\}$. The network objective is to solve the following constrained optimization problem:

$$\text{minimize} \quad \sum_{i=1}^{m} f_i(x)$$
$$\text{subject to} \quad x \in X, \tag{2.1}$$

where $X \subseteq \Re^n$ is a constraint set and $f_i : X \to \Re$ for all i. Related to the problem we use the following notation

$$f(x) = \sum_{i=1}^{m} f_i(x), \qquad f^* = \min_{x \in X} f(x), \qquad X^* = \{x \in X : f(x) = f^*\}.$$

We are interested in the case when the problem in (2.1) is convex. Specifically, we assume that the following assumption holds.

ASSUMPTION 1. *The functions f_i and the set X are such that*
(a) The set X is closed and convex.
(b) The functions f_i, $i \in V$ are defined and convex over an open set that contains the set X.

The function f_i is known only partially to agent i in the sense that the agent can only obtain a noisy estimate of the function subgradient. The goal is to solve problem (2.1) using an algorithm that is distributed and local.[1]

We make no assumption on the differentiability of the functions f_i. At points where the gradient does not exist, we use the notion of subgradients. A vector ∇f_i is a subgradient of f_i at a point $x \in \text{dom} f$ if the following relation holds

$$\nabla f_i(x)^T (y - x) \leq f_i(y) - f_i(x) \qquad \text{for all } y \in \text{dom} f. \tag{2.2}$$

Since the set X is contained in an open set over which the functions are defined and convex, a subgradient of f_i exists at any point of the set X (see [1] or [27]).

2.2. Algorithm. To solve the problem in (2.1) with its inherent decentralized information access, we consider an iterative subgradient method. The iterations are distributed accordingly among the agents, whereby each agent i is minimizing its convex objective f_i over the set X and locally exchanging the iterates with its neighbors.

Let $w_{i,k}$ be the iterate with agent i at the end of iteration k. At the beginning of iteration $k + 1$, agent i receives the current iterate of a subset of the agents. Then, agent i computes a weighted average of these iterates and adjusts this average along the negative subgradient direction of f_i, which is computed with stochastic errors. The adjusted iterate is then projected onto the constraint set X. Mathematically, each agent i generates its iterate sequence $\{w_{i,k}\}$ according to the following relation:

$$w_{i,k+1} = P_X \left[v_{i,k} - \alpha_{k+1} \left(\nabla f_i \left(v_{i,k} \right) + \epsilon_{i,k+1} \right) \right], \tag{2.3}$$

starting with some initial iterate $w_{i,0} \in X$. Here, $\nabla f_i \left(v_{i,k} \right)$ denotes the subgradient of f_i at $v_{i,k}$ and $\epsilon_{i,k+1}$ is the stochastic error in the subgradient evaluation. The scalar

---

[1]See [25, 26] for wireless network applications that can be cast in this framework.

$\alpha_{k+1} > 0$ is the stepsize and P_X denotes the Euclidean projection onto the set X. The vector $v_{i,k}$ is the weighted average computed by agent i and is given by

$$v_{i,k} = \sum_{j \in N_i(k+1)} a_{i,j}(k+1)w_{j,k}, \qquad (2.4)$$

where $N_i(k+1)$ denotes the set of agents whose current iterates are available to agent i in the $(k+1)$-st iteration. We assume that $i \in N_i(k+1)$ for all agents and at all times k. The scalars $a_{i,j}(k+1)$ are the non-negative weights that agent i assigns to agent j's iterate. We will find it convenient to define $a_{i,j}(k+1)$ as 0 for $j \notin N_i(k+1)$ and rewrite (2.4) as

$$v_{i,k} = \sum_{j=1}^{m} a_{i,j}(k+1)w_{j,k}. \qquad (2.5)$$

This is a "consensus"-based step ensuring that, in a long run, the information of each f_i reaches every agent with the same frequency, directly or through a sequence of local communications. Due to this, the iterates $w_{j,k}$ become eventually "the same" for all j and for large enough k. The update step in (2.3) is just a subgradient iteration for minimizing f_i over X taken after the "consensus"-based step.

2.3. Additional assumptions. In addition to Assumption 1, we make some assumptions on the inter-agent exchange model and the weights. The first assumption requires the agents to communicate sufficiently often so that all the component functions, directly or indirectly, influence the iterate sequence of any agent. Recall that we defined $N_i(k + 1)$ as the set of agents that agent i communicates with in iteration $k + 1$. Define (V, E_{k+1}) to be the graph with edges

$$E_{k+1} = \{(j,i) : j \in N_i(k+1), i \in V\}.$$

ASSUMPTION 2. *There exists a scalar Q such that the graph $(V, \cup_{l=1,\dots,Q} E_{k+l})$ is strongly connected for all k.*

It is also important that the influence of the functions f_i is "equal" in a long run so that the sum of the component functions is minimized rather than a weighted sum of them. The influence of a component f_j on the iterates of agent i depends on the weights that agent i uses. To ensure equal influence, we make the following assumption on the weights.

ASSUMPTION 3. *For $i \in V$ and all k,*
(a) $a_{i,j}(k + 1) \geq 0$, and $a_{i,j}(k + 1) = 0$ when $j \notin N_i(k + 1)$,
(b) $\sum_{j=1}^{m} a_{i,j}(k + 1) = 1$,
(c) There exists a scalar η, $0 < \eta < 1$, such that $a_{i,j}(k+1) \geq \eta$ when $j \in N_i(k+1)$,
(d) $\sum_{i=1}^{m} a_{i,j}(k + 1) = 1$.

Assumptions 3a and 3b state that each agent calculates a weighted average of all the iterates it has access to. Assumption 3c ensures that each agent gives a sufficient weight to its current iterate and all the iterates it receives.[2] Assumption 3d, together with Assumption 2, as we will see later, ensures that all the agents are equally influential in the long run. In other words, Assumption 3d is crucial to ensure that $\sum_{i=1}^{m} f_i$ is minimized as opposed to a weighted sum of the functions f_i with non-equal weights. To satisfy Assumption 3d, the agents need to coordinate their weights. Some coordination schemes are discussed in [19, 25].

---

[2] The agents need not be aware of the common bound η.

3. Preliminaries. In this section, we state some results for future reference.

3.1. Euclidean norm inequalities. For any vectors $v_1, \ldots, v_M \in \Re^n$, we have

$$\sum_{i=1}^{M} \left\| v_i - \frac{1}{M} \sum_{j=1}^{M} v_j \right\|^2 \leq \sum_{i=1}^{M} \|v_i - x\|^2 \qquad \text{for any } x \in \Re^n. \tag{3.1}$$

The preceding relation states that the average of a finite set of vectors minimizes the sum of distances between each vector and any vector in $\Re^n$, which can be verified using the first-order optimality conditions.

Both the Euclidean norm and its square are convex functions, i.e., for any vectors $v_1, \ldots, v_M \in \Re^n$ and nonnegative scalars $\beta_1, \ldots, \beta_M$ such that $\sum_{i=1}^{M} \beta_i = 1$, we have

$$\left\| \sum_{i=1}^{M} \beta_i v_i \right\| \leq \sum_{i=1}^{M} \beta_i \|v_i\|, \tag{3.2}$$

$$\left\| \sum_{i=1}^{M} \beta_i v_i \right\|^2 \leq \sum_{i=1}^{M} \beta_i \|v_i\|^2. \tag{3.3}$$

The following inequality is the well-known[3] *non-expansive* property of the Euclidean projection onto a nonempty, closed and convex set X,

$$\|P_X[x] - P_X[y]\| \leq \|x - y\| \qquad \text{for all } x, y \in \Re^n. \tag{3.4}$$

3.2. Scalar sequences. For a scalar β and a scalar sequence $\{\gamma_k\}$, we consider the "convolution" sequence $\sum_{\ell=0}^{k} \beta^{k-\ell} \gamma_\ell = \beta^k \gamma_0 + \beta^{k-1} \gamma_1 + \cdots + \beta \gamma_{k-1} + \gamma_k$. We have the following result.

LEMMA 3.1. *Let $\{\gamma_k\}$ be a scalar sequence.*

(a) If $\lim_{k \to \infty} \gamma_k = \gamma$ and $0 < \beta < 1$, then $\lim_{k \to \infty} \sum_{\ell=0}^{k} \beta^{k-\ell} \gamma_\ell = \frac{\gamma}{1-\beta}$.

(b) If $\gamma_k \geq 0$ for all k, $\sum_k \gamma_k < \infty$ and $0 < \beta < 1$, then $\sum_{k=0}^{\infty} \left(\sum_{\ell=0}^{k} \beta^{k-\ell} \gamma_\ell \right) < \infty$.

(c) If $\limsup_{k \to \infty} \gamma_k = \gamma$ and $\{\zeta_k\}$ is a positive scalar sequence with $\sum_{k=1}^{\infty} \zeta_k = \infty$, then $\limsup_{K \to \infty} \frac{\sum_{k=0}^{K} \gamma_k \zeta_k}{\sum_{k=0}^{K} \zeta_k} \leq \gamma$. In addition, if $\liminf_{k \to \infty} \gamma_k = \gamma$, then $\lim_{K \to \infty} \frac{\sum_{k=0}^{K} \gamma_k \zeta_k}{\sum_{k=0}^{K} \zeta_k} = \gamma$.

Proof. (a) Let $\epsilon > 0$ be arbitrary. Since $\gamma_k \to \gamma$ and for all k, there is an index K such that $|\gamma_k - \gamma| \leq \epsilon$ for all $k \geq K$. For all $k \geq K+1$, we have

$$\sum_{\ell=0}^{k} \beta^{k-\ell} \gamma_\ell = \sum_{\ell=0}^{K} \beta^{k-\ell} \gamma_\ell + \sum_{\ell=K+1}^{k} \beta^{k-\ell} \gamma_\ell \leq \max_{0 \leq t \leq K} \gamma_t \sum_{\ell=0}^{K} \beta^{k-\ell} + (\gamma + \epsilon) \sum_{\ell=K+1}^{k} \beta^{k-\ell}.$$

Since $\sum_{\ell=K+1}^{k} \beta^{k-\ell} \leq \frac{1}{1-\beta}$ and

$$\sum_{\ell=0}^{K} \beta^{k-\ell} = \beta^k + \cdots + \beta^{k-K} = \beta^{k-K}(1 + \cdots + \beta^K) \leq \frac{\beta^{k-K}}{1-\beta},$$

---

[3]See for example [1], Proposition 2.2.1.

it follows that for all $k \geq K + 1$,

$$\sum_{\ell=0}^{k} \beta^{k-\ell} \gamma_\ell \leq \left(\max_{0 \leq t \leq K} \gamma_t \right) \frac{\beta^{k-K}}{1-\beta} + \frac{\gamma + \epsilon}{1-\beta}.$$

Therefore,

$$\limsup_{k \to \infty} \sum_{\ell=0}^{k} \beta^{k-\ell} \gamma_\ell \leq \frac{\gamma + \epsilon}{1-\beta}.$$

Since ϵ is arbitrary, we conclude that $\limsup_{k \to \infty} \sum_{\ell=0}^{k} \beta^{k-\ell} \gamma_\ell \leq \frac{\gamma}{1-\beta}$.

Similarly, we have

$$\sum_{\ell=0}^{k} \beta^{k-\ell} \gamma_\ell \geq \min_{0 \leq t \leq K} \gamma_t \sum_{\ell=0}^{K} \beta^{k-\ell} + (\gamma - \epsilon) \sum_{\ell=K+1}^{k} \beta^{k-\ell}.$$

Thus,

$$\liminf_{k \to \infty} \sum_{\ell=0}^{k} \beta^{k-\ell} \gamma_\ell \geq \liminf_{k \to \infty} \left(\min_{0 \leq t \leq K} \gamma_t \sum_{\ell=0}^{K} \beta^{k-\ell} + (\gamma - \epsilon) \sum_{\ell=K+1}^{k} \beta^{k-\ell} \right).$$

Since $\sum_{\ell=0}^{K} \beta^{k-\ell} \geq \beta^{k-K}$ and $\sum_{\ell=K+1}^{k} \beta^{k-\ell} = \sum_{s=0}^{k-(K+1)} \beta^s$, which tends to $1/(1-\beta)$ as $k \to \infty$, it follows that

$$\liminf_{k \to \infty} \sum_{\ell=0}^{k} \beta^{k-\ell} \gamma_\ell \geq \left(\min_{0 \leq t \leq K} \gamma_t \right) \lim_{k \to \infty} \beta^{k-K} + (\gamma - \epsilon) \lim_{k \to \infty} \sum_{s=0}^{k-(K+1)} \beta^s = \frac{\gamma - \epsilon}{1-\beta}.$$

Since ϵ is arbitrary, we have $\liminf_{k \to \infty} \sum_{\ell=0}^{k} \beta^{k-\ell} \gamma_\ell \geq \frac{\gamma}{1-\beta}$. This and the relation $\limsup_{k \to \infty} \sum_{\ell=0}^{k} \beta^{k-\ell} \gamma_\ell \leq \frac{\gamma}{1-\beta}$, imply

$$\lim_{k \to \infty} \sum_{\ell=0}^{k} \beta^{k-\ell} \gamma_\ell = \frac{\gamma}{1-\beta}.$$

(b) Let $\sum_{k=0}^{\infty} \gamma_k < \infty$. For any integer $M \geq 1$, we have

$$\sum_{k=0}^{M} \left(\sum_{\ell=0}^{k} \beta^{k-\ell} \gamma_\ell \right) = \sum_{\ell=0}^{M} \gamma_\ell \sum_{t=0}^{M-\ell} \beta^t \leq \sum_{\ell=0}^{M} \gamma_\ell \frac{1}{1-\beta},$$

implying that

$$\sum_{k=0}^{\infty} \left(\sum_{\ell=0}^{k} \beta^{k-\ell} \gamma_\ell \right) \leq \frac{1}{1-\beta} \sum_{\ell=0}^{\infty} \gamma_\ell < \infty.$$

(c) Since $\limsup_{k \to \infty} \gamma_k = \gamma$, for every $\epsilon > 0$ there is a large enough K such that $\gamma_k \leq \gamma + \epsilon$ for all $k \geq K$. Thus, for any $M > K$,

$$\frac{\sum_{k=0}^{M} \gamma_k \zeta_k}{\sum_{k=0}^{M} \zeta_k} = \frac{\sum_{k=0}^{K} \gamma_k \zeta_k}{\sum_{k=0}^{M} \zeta_k} + \frac{\sum_{k=K+1}^{M} \gamma_k \zeta_k}{\sum_{k=0}^{M} \zeta_k} \leq \frac{\sum_{k=0}^{K} \gamma_k \zeta_k}{\sum_{k=0}^{M} \zeta_k} + (\gamma + \epsilon) \frac{\sum_{k=K+1}^{M} \zeta_k}{\sum_{k=0}^{M} \zeta_k}.$$

By letting $M \to \infty$ and using $\sum_k \zeta_k = \infty$, we see that $\limsup_{M \to \infty} \frac{\sum_{k=0}^M \gamma_k \zeta_k}{\sum_{k=0}^M \zeta_k} \leq \gamma + \epsilon$, and since ϵ is arbitrary, the result for the limit superior follows.

Analogously, if $\liminf_{k \to \infty} \gamma_k = \gamma$, then for every $\epsilon > 0$ there is a large enough K such that $\gamma_k \geq \gamma - \epsilon$ for all $k \geq K$. Thus, for any $M > K$,

$$\frac{\sum_{k=0}^M \gamma_k \zeta_k}{\sum_{k=0}^M \zeta_k} = \frac{\sum_{k=0}^K \gamma_k \zeta_k}{\sum_{k=0}^M \zeta_k} + \frac{\sum_{k=K+1}^M \gamma_k \zeta_k}{\sum_{k=0}^M \zeta_k} \geq \frac{\sum_{k=0}^K \gamma_k \zeta_k}{\sum_{k=0}^M \zeta_k} + (\gamma - \epsilon) \frac{\sum_{k=K+1}^M \zeta_k}{\sum_{k=0}^M \zeta_k}.$$

Letting $M \to \infty$ and using $\sum_k \zeta_k = \infty$, we obtain $\liminf_{M \to \infty} \frac{\sum_{k=0}^M \gamma_k \zeta_k}{\sum_{k=0}^M \zeta_k} \geq \gamma - \epsilon$. Since $\epsilon > 0$ is arbitrary, we have $\liminf_{M \to \infty} \frac{\sum_{k=0}^M \gamma_k \zeta_k}{\sum_{k=0}^M \zeta_k} \geq \gamma$. This relation and the relation for the limit superior yield $\lim_{M \to \infty} \frac{\sum_{k=0}^M \gamma_k \zeta_k}{\sum_{k=0}^M \zeta_k} = \gamma$ when $\gamma_k \to \gamma$. $\square$

3.3. Matrix convergence. Let $A(k)$ be the matrix with (i,j)-th entry equal to $a_{i,j}(k)$. As a consequence of Assumptions 3a, 3b and 3d, the matrix $A(k)$ is doubly stochastic[4]. Define, for all k, s with $k \geq s$,

$$\Phi(k, s) = A(k)A(k-1) \cdots A(s+1). \tag{3.5}$$

We next state a result from [18] (Corollary 1) on the convergence properties of the matrix $\Phi(k, s)$. Let $[\Phi(k, s)]_{i,j}$ denote the (i,j)-th entry of the matrix $\Phi(k, s)$, and let $e \in \Re^m$ be the column vector with all entries equal to 1.

LEMMA 3.2. *Let Assumptions 2 and 3 hold. Then*
1. *$\lim_{k \to \infty} \Phi(k, s) = \frac{1}{m} ee^T$ for all s.*
2. *Further, the convergence is geometric and the rate of convergence is given by*

$$\left| [\Phi(k, s)]_{i,j} - \frac{1}{m} \right| \leq \theta \beta^{k-s},$$

where

$$\theta = \left(1 - \frac{\eta}{4m^2}\right)^{-2} \qquad \beta = \left(1 - \frac{\eta}{4m^2}\right)^{\frac{1}{Q}}.$$

3.4. Stochastic convergence. We next state some results that deal with the convergence of a sequence of random vectors. The first result is the well known Fatou's lemma [4].

LEMMA 3.3. *Let $\{X_i\}$ be a sequence of non-negative random variables. Then*

$$\mathsf{E}\left[\liminf_{n \to \infty} X_n\right] \leq \liminf_{n \to \infty} \mathsf{E}[X_n].$$

The next result is due to Robbins and Siegmund (Lemma 11, Chapter 2.2, [22]).

THEOREM 3.4. *Let $\{B_k\}, \{D_k\}$, and $\{H_k\}$ be non-negative random sequences and let $\{\zeta_k\}$ be a deterministic nonnegative scalar sequence. Let G_k be the σ-algebra generated by $B_1, \ldots, B_k, D_1, \ldots, D_k, H_1, \ldots, H_k$. Suppose that $\sum_k \zeta_k < \infty$,*

$$\mathsf{E}[B_{k+1} \mid G_k] \leq (1 + \zeta_k)B_k - D_k + H_k \qquad \text{for all } k, \tag{3.6}$$

and $\sum_k H_k < \infty$ with probability 1. Then, the sequence $\{B_k\}$ converges to a non-negative random variable and $\sum_k D_k < \infty$ with probability 1, and in mean.

---

[4]The sum of its entries in every row and in every column is equal to 1.

4. Basic relations. In this section, we derive two basic relations that form the basis for the analysis in this paper. The first of them deals with the disagreements among the agents, and the second deals with the agent iterate sequences.

4.1. Disagreement Estimate. The agent disagreements are typically thought of as the norms $\|w_{i,k} - w_{j,k}\|$ of the differences between the iterates $w_{i,k}$ and $w_{j,k}$ generated by different agents according to (2.3)–(2.4). Alternatively, the agent disagreements can be measured with respect to a reference sequence, which we adopt here. In particular, we study the behavior of $\|y_k - w_{i,k}\|$, where $\{y_k\}$ is the auxiliary vector sequence defined by

$$y_k = \frac{1}{m} \sum_{i=1}^{m} w_{i,k} \qquad \text{for all } k. \tag{4.1}$$

In the next lemma, we provide a basic estimate for $\|y_k - w_{j,k}\|$. The rate of convergence result from Lemma 3.2 plays a crucial role in obtaining this estimate.

LEMMA 4.1. *Let Assumptions 1a, 2, and 3 hold. Assume that the subgradients of f_i are uniformly bounded over the set X, i.e., there are scalars C_i such that*

$$\|\nabla f_i(x)\| \le C_i \qquad \text{for all } x \in X \text{ and all } i \in V.$$

Then, for all $j \in V$ and $k \ge 0$,

$$\|y_{k+1} - w_{j,k+1}\| \le m\theta\beta^{k+1} \max_{i \in V} \|w_{i,0}\| + \theta \sum_{\ell=1}^{k} \alpha_\ell \beta^{k+1-\ell} \sum_{i=1}^{m} (C_i + \|\epsilon_{i,\ell}\|)$$

$$+ \frac{\alpha_{k+1}}{m} \sum_{i=1}^{m} (C_i + \|\epsilon_{i,k+1}\|) + \alpha_{k+1}(C_j + \|\epsilon_{j,k+1}\|).$$

Proof. Define for all $i \in V$ and all k,

$$p_{i,k+1} = w_{i,k+1} - \sum_{j=1}^{m} a_{i,j}(k+1)w_{j,k}. \tag{4.2}$$

Using the matrices $\Phi(k,s)$ defined in (3.5) we can write

$$w_{j,k+1} = \sum_{i=1}^{m} [\Phi(k+1,0)]_{j,i} w_{i,0} + p_{j,k+1} + \sum_{\ell=1}^{k} \left(\sum_{i=1}^{m} [\Phi(k+1,\ell)]_{j,i} p_{i,\ell} \right). \tag{4.3}$$

Using (4.2), we can also rewrite y_k, defined in (4.1), as follows

$$y_{k+1} = \frac{1}{m} \left(\sum_{i=1}^{m} \sum_{j=1}^{m} a_{i,j}(k+1)w_{j,k} + \sum_{i=1}^{m} p_{i,k+1} \right)$$

$$= \frac{1}{m} \left(\sum_{j=1}^{m} \left(\sum_{i=1}^{m} a_{i,j}(k+1) \right) w_{j,k} + \sum_{i=1}^{m} p_{i,k+1} \right).$$

In the view of the doubly stochasticity of the weights, we have $\sum_{i=1}^{m} a_{i,j}(k+1) = 1$, implying that

$$y_{k+1} = \frac{1}{m} \left(\sum_{j=1}^{m} w_{j,k} + \sum_{i=1}^{m} p_{i,k+1} \right) = y_k + \frac{1}{m} \sum_{i=1}^{m} p_{i,k+1}.$$

Therefore

$$y_{k+1} = y_0 + \frac{1}{m} \sum_{\ell=1}^{k+1} \sum_{i=1}^{m} p_{i,\ell} = \frac{1}{m} \sum_{i=1}^{m} w_{i,0} + \frac{1}{m} \sum_{\ell=1}^{k+1} \sum_{i=1}^{m} p_{i,\ell}. \qquad (4.4)$$

Substituting for y_{k+1} from (4.4) and for $w_{j,k+1}$ from (4.3), we obtain

$$
\begin{aligned}
\|y_{k+1} - w_{j,k+1}\| = &\left\| \frac{1}{m} \sum_{i=1}^{m} w_{i,0} + \frac{1}{m} \sum_{\ell=1}^{k+1} \sum_{i=1}^{m} p_{i,\ell} \right. \\
&- \left. \left(\sum_{i=1}^{m} [\Phi(k+1,0)]_{j,i} w_{i,0} + p_{j,k+1} + \sum_{\ell=1}^{k} \sum_{i=1}^{m} [\Phi(k+1,\ell)]_{j,i} p_{i,\ell} \right) \right\| \\
= &\left\| \sum_{i=1}^{m} \left(\frac{1}{m} - [\Phi(k+1,0)]_{j,i} \right) w_{i,0} \right. \\
&+ \left. \sum_{\ell=1}^{k} \sum_{i=1}^{m} \left(\frac{1}{m} - [\Phi(k+1,\ell)]_{j,i} \right) p_{i,\ell} + \left(\frac{1}{m} \sum_{i=1}^{m} p_{i,k+1} - p_{j,k+1} \right) \right\|.
\end{aligned}
$$

Therefore, for all $j \in V$ and all k,

$$
\begin{aligned}
\|y_{k+1} - w_{j,k+1}\| \leq &\sum_{i=1}^{m} \left| \frac{1}{m} - [\Phi(k+1,0)]_{j,i} \right| \|w_{i,0}\| \\
&+ \sum_{\ell=1}^{k} \sum_{i=1}^{m} \left| \frac{1}{m} - [\Phi(k+1,\ell)]_{j,i} \right| \|p_{i,\ell}\| + \frac{1}{m} \sum_{i=1}^{m} \|p_{i,k+1}\| + \|p_{j,k+1}\|.
\end{aligned}
$$

We can bound $\|w_{i,0}\| \leq \max_{i \in V} \|w_{i,0}\|$. Further, we can use the rate of convergence result from Lemma 3.2 to bound $\left| \frac{1}{m} - [\Phi(k,\ell)]_{j,i} \right|$. We obtain

$$
\begin{aligned}
\|y_{k+1} - w_{j,k+1}\| \leq &m\theta\beta^{k+1} \max_{i \in V} \|w_{i,0}\| + \theta \sum_{\ell=1}^{k} \beta^{k+1-\ell} \sum_{i=1}^{m} \|p_{i,\ell}\| \\
&+ \frac{1}{m} \sum_{i=1}^{m} \|p_{i,k+1}\| + \|p_{j,k+1}\|. \qquad (4.5)
\end{aligned}
$$

We next estimate the norms of the vectors $\|p_{i,k}\|$ for any k. From the definition of $p_{i,k+1}$ in (4.2) and the definition of the vector $v_{i,k}$ in (2.4), we have $p_{i,k+1} = w_{i,k+1} - v_{i,k}$. Note that, being a convex combination of vectors $w_{j,k}$ in the convex set X, the vector $v_{i,k}$ is in the set X. By the definition of the iterate $w_{i,k+1}$ in (2.3) and the non-expansive property of the Euclidean projection in (3.4), we have

$$
\begin{aligned}
\|p_{i,k+1}\| &= \|P_X [v_{i,k} - \alpha_{k+1} (\nabla f_i(v_{i,k}) + \epsilon_{i,k+1})] - v_{i,k}\| \\
&\leq \alpha_{k+1} \|\nabla f_i(v_{i,k}) + \epsilon_{i,k+1}\| \\
&\leq \alpha_{k+1} (C_i + \|\epsilon_{i,k+1}\|).
\end{aligned}
$$

In the last step we have used the subgradient boundedness. By substituting the preceding relation in (4.5), we obtain the desired relation. $\square$

4.2. Iterate Relation. Here, we derive a relation for the distances $\|v_{i,k+1} - z\|$ and the function value differences $f(y_k) - f(z)$ for an arbitrary $z \in X$. This relation

together with Lemma 4.1 provides the basis for our subsequent convergence analysis. In what follows, recall that $f = \sum_{i=1}^{m} f_i$.

LEMMA 4.2. *Let Assumptions 1, 2, and 3 hold. Assume that the subgradients of f_i are uniformly bounded over the set X, i.e., there are scalars C_i such that*

$$\|\nabla f_i(x)\| \le C_i \qquad \text{for all } x \in X \text{ and all } i \in V.$$

Then, for any $z \in X$ and all k,

$$\sum_{i=1}^{m} \|v_{i,k+1} - z\|^2 \le \sum_{i=1}^{m} \|v_{i,k} - z\|^2 - 2\alpha_{k+1} \left(f(y_k) - f(z) \right)$$

$$+2\alpha_{k+1} \left(\max_{i \in V} C_i \right) \sum_{j=1}^{m} \|y_k - w_{j,k}\|$$

$$-2\alpha_{k+1} \sum_{i=1}^{m} \epsilon_{i,k+1}^T (v_{i,k} - z) + \alpha_{k+1}^2 \sum_{i=1}^{m} \left(C_i + \|\epsilon_{i,k+1}\| \right)^2.$$

Proof. Using the Euclidean projection property in (3.4), from the definition of the iterate $w_{i,k+1}$ in (2.3), we have for any $z \in X$ and all k,

$$\|w_{i,k+1} - z\|^2 = \|P_X \left[v_{i,k} - \alpha_{k+1} \left(\nabla f_i(v_{i,k}) + \epsilon_{i,k+1} \right) \right] - z\|^2$$

$$\le \|v_{i,k} - z\|^2 - 2\alpha_{k+1} \nabla f_i(v_{i,k})^T (v_{i,k} - z) - 2\alpha_{k+1} \epsilon_{i,k+1}^T (v_{i,k} - z)$$

$$+ \alpha_{k+1}^2 \|\nabla f_i(v_{i,k}) + \epsilon_{i,k+1}\|^2.$$

By using the subgradient inequality in (2.2) to bound the second term, we obtain

$$\|w_{i,k+1} - z\|^2 \le \|v_{i,k} - z\|^2 - 2\alpha_{k+1} \left(f_i(v_{i,k}) - f_i(z) \right)$$

$$- 2\alpha_{k+1} \epsilon_{i,k+1}^T (v_{i,k} - z) + \alpha_{k+1}^2 \|\nabla f_i(v_{i,k}) + \epsilon_{i,k+1}\|^2. \quad (4.6)$$

Note that by the convexity of the squared norm [cf. Eq. (3.3)], we have

$$\sum_{i=1}^{m} \|v_{i,k+1} - z\|^2 = \sum_{i=1}^{m} \left\| \sum_{j=1}^{m} a_{i,j}(k+2) w_{j,k+1} - z \right\|^2 \le \sum_{i=1}^{m} \sum_{j=1}^{m} a_{i,j}(k+2) \|w_{j,k+1} - z\|^2.$$

In view of Assumption 3, we have $\sum_{i=1}^{m} a_{i,j}(k+2) = 1$ for all j and k, implying that

$$\sum_{i=1}^{m} \|v_{i,k+1} - z\|^2 \le \sum_{j=1}^{m} \|w_{j,k+1} - z\|^2.$$

By summing the relations in (4.6) over all $i \in V$ and by using the preceding relation, we obtain

$$\sum_{i=1}^{m} \|v_{i,k+1} - z\|^2 \le \sum_{i=1}^{m} \|v_{i,k} - z\|^2 - 2\alpha_{k+1} \sum_{i=1}^{m} \left(f_i(v_{i,k}) - f_i(z) \right)$$

$$-2\alpha_{k+1} \sum_{i=1}^{m} \epsilon_{i,k+1}^T (v_{i,k} - z) + \alpha_{k+1}^2 \sum_{i=1}^{m} \|\nabla f_i(v_{i,k}) + \epsilon_{i,k+1}\|^2. (4.7)$$

From (2.2) we have

$$f_i(v_{i,k}) - f_i(z) \geq (f_i(v_{i,k}) - f_i(y_k)) + (f_i(y_k) - f_i(z))$$
$$\geq - \|\nabla f_i(v_{i,k})\| \|y_k - v_{i,k}\| + (f_i(y_k) - f_i(z)). \qquad (4.8)$$

Recall that $v_{i,k} = \sum_{j=1}^m a_{i,j}(k+1)w_{j,k}$ [cf. (2.5)]. Substituting for $v_{i,k}$ and using the convexity of the norm [cf. (3.2)], from (4.8) we obtain

$$\sum_{i=1}^m f_i(v_{i,k}) - f_i(z) \geq - \sum_{i=1}^m \|\nabla f_i(v_{i,k})\| \|y_k - v_{i,k}\| + (f(y_k) - f(z))$$

$$\geq - \sum_{i=1}^m \|\nabla f_i(v_{i,k})\| \left\| y_k - \sum_{j=1}^m a_{i,j}(k+1)w_{j,k} \right\| + (f(y_k) - f(z))$$

$$\geq - \sum_{i=1}^m \|\nabla f_i(v_{i,k})\| \sum_{j=1}^m a_{i,j}(k+1) \|y_k - w_{j,k}\| + (f(y_k) - f(z))$$

$$\geq - \left(\max_{i \in V} \|\nabla f_i(v_{i,k})\| \right) \sum_{j=1}^m \left(\sum_{i=1}^m a_{i,j}(k+1) \right) \|y_k - w_{j,k}\|$$
$$+ (f(y_k) - f(z))$$

$$= - \left(\max_{i \in V} \|\nabla f_i(v_{i,k})\| \right) \sum_{j=1}^m \|y_k - w_{j,k}\| + (f(y_k) - f(z)).$$

By using the preceding estimate in relation (4.7), we have

$$\sum_{i=1}^m \|v_{i,k+1} - z\|^2 \leq \sum_{i=1}^m \|v_{i,k} - z\|^2 - 2\alpha_{k+1}(f(y_k) - f(z))$$

$$+ 2\alpha_{k+1} \left(\max_{i \in V} \|\nabla f_i(v_{i,k})\| \right) \sum_{j=1}^m \|y_k - w_{j,k}\|$$

$$- 2\alpha_{k+1} \sum_{i=1}^m \epsilon_{i,k+1}^T (v_{i,k} - z) + \alpha_{k+1}^2 \sum_{i=1}^m \|\nabla f_i(v_{i,k}) + \epsilon_{i,k+1}\|^2.$$

The result follows by using the subgradient norm boundedness, $\|\nabla f_i(v_{i,k})\| \leq C_i$ for all k and i. $\square$

5. Convergence in mean. Here, we study the behavior of the iterates generated by the algorithm, under the assumption that the errors have bounded norms in mean square. In particular, we assume the following.

ASSUMPTION 4. *The subgradient errors are uniformly bounded in mean square, i.e, there are scalars $\bar{\nu}_i$ such that*

$$\mathsf{E}[\|\epsilon_{i,k+1}\|^2] \leq \bar{\nu}_i^2 \qquad \text{for all } i \in V \text{ and all } k.$$

Using this assumption, we provide a bound on the expected disagreement $\mathsf{E}[\|w_{i,k} - y_k\|]$ for nondiminishing stepsize. We later use this bound to provide an estimate for the algorithm's performance in mean. The bound is provided in the following theorem.

THEOREM 5.1. *Let Assumptions 1a, 2, 3 and 4 hold. Also, let the subgradients of each f_i be uniformly bounded over X, i.e., for each $i \in V$ there is C_i such that*

$$\|\nabla f_i(x)\| \leq C_i \qquad \text{for all } x \in X.$$

If the stepsize $\{\alpha_k\}$ is such that $\lim_{k\to\infty} \alpha_k = \alpha$ for some $\alpha \geq 0$, then for all $j \in V$,

$$\limsup_{k\to\infty} \mathsf{E}[\|y_{k+1} - w_{j,k+1}\|] \leq \alpha \max_{i\in V}\{C_i + \bar{\nu}_i\} \left(2 + \frac{m\theta\beta}{1-\beta}\right).$$

Proof. The conditions of Lemma 4.1 are satisfied. Taking the expectation in the relation of Lemma 4.1 and using the inequality $\mathsf{E}[\|\epsilon_{i,k}\|] \leq \sqrt{\mathsf{E}[\|\epsilon_{i,k}\|^2]} = \bar{\nu}_i$, we obtain for all $j \in V$ and all k,

$$\mathsf{E}[\|y_{k+1} - w_{j,k+1}\|] \leq m\theta\beta^{k+1} \max_{i\in V} \|w_{i,0}\| + m\theta\beta \max_{i\in V}\{C_i + \bar{\nu}_i\} \sum_{\ell=1}^{k} \beta^{k-\ell}\alpha_\ell$$
$$+ 2\alpha_{k+1}\max_{i\in V}\{C_i + \bar{\nu}_i\}. \tag{5.1}$$

Since $\lim_{k\to\infty} \alpha_k = \alpha$, by Lemma 3.1(a) we have $\lim_{k\to\infty} \sum_{\ell=1}^{k} \beta^{k-\ell}\alpha_\ell = \frac{\alpha}{1-\beta}$. Using this relation and $\lim_{k\to\infty} \alpha_k = \alpha$, we obtain the result by taking the limit superior in (5.1) as $k \to \infty$. $\square$

When the stepsize is diminishing (i.e., $\alpha = 0$), the result of Theorem 5.1 implies that the expected disagreements $\mathsf{E}[\|y_{k+1} - w_{j,k+1}\|]$ converge to 0 for all j. Thus, there is an asymptotic consensus in mean. We formally state this as a corollary.

COROLLARY 5.2. *Let the conditions of Theorem 5.1 hold with $\alpha = 0$. Then $\lim_{k\to\infty} \mathsf{E}[\|w_{j,k} - y_k\|] = 0$ for all $j \in V$.*

We next obtain bounds on the performance of the algorithm. We make the additional assumption that the set X is bounded. Thus, the subgradients of each f_i are also bounded (see [1], Proposition 4.2.3).

Note that, under Assumption 4, by Jensen's inequality we have $\|\mathsf{E}[\epsilon_{i,k+1}]\| \leq \bar{\nu}_i$. Therefore, under Assumption 4,

$$\limsup_{k\to\infty} \|\mathsf{E}[\epsilon_{i,k+1}]\| \leq \bar{\nu}_i \qquad \text{for all } i \in V. \tag{5.2}$$

We have used this relation in our analysis of the agent disagreements in Theorem 5.1. Using this relation, we obtain special results for the cases when the errors are zero mean or when their mean is diminishing, i.e., the cases $\mathsf{E}[\epsilon_{i,k+1}] = 0$ for all i, k, or $\limsup_{k\to\infty} \|\mathsf{E}[\epsilon_{i,k+1}]\| = 0$ for all i.

THEOREM 5.3. *Let Assumptions 1, 2, 3 and 4 hold. Assume that the set X is bounded. Let $\lim_{k\to\infty} \alpha_k = \alpha$ with $\alpha \geq 0$. If $\alpha = 0$, also assume that $\sum_k \alpha_k = \infty$. Then, for all $j \in V$,*

$$\liminf_{k\to\infty} \mathsf{E}[f(w_{j,k})] \leq f^* + \max_{x,y\in X} \|x - y\| \sum_{i=1}^{m} \bar{\mu}_i + m\alpha \left(\max_{i\in V}\{C_i + \bar{\nu}_i\}\right)^2 \left(\frac{9}{2} + \frac{2m\theta\beta}{1-\beta}\right),$$

where $\bar{\mu}_i = \limsup_{k\to\infty} \|\mathsf{E}[\epsilon_{i,k+1}]\|$ and C_i is an upper-bound on the subgradient norms of f_i over the set X.

Proof. Under Assumption 4, the limit superiors $\bar{\mu}_i = \limsup_{k\to\infty} \|\mathsf{E}[\epsilon_{i,k+1}]\|$ are finite [cf. Eq. (5.2)]. Since the set X is bounded the subgradients of f_i over the set X are also bounded for each $i \in V$; hence, the bounds C_i, $i \in V$ on subgradient norms exist. Thus, the conditions of Lemma 4.2 are satisfied. Further, by Assumption 1, the set X is contained in the interior of the domain of f, over which the function is continuous (by convexity; see [27]). Thus, the set X is compact and f is continuous

over X, implying that the optimal set X^* is nonempty. Let $x^* \in X^*$, and let $y = x^*$ in Lemma 4.2. We have, for all k,

$$\sum_{i=1}^m \|v_{i,k+1} - x^*\|^2 \le \sum_{i=1}^m \|v_{i,k} - x^*\|^2 - 2\alpha_{k+1}\left(f(y_k) - f^*\right)$$

$$+2\alpha_{k+1}\left(\max_{i \in V} C_i\right)\sum_{j=1}^m \|y_k - w_{j,k}\|$$

$$-2\alpha_{k+1}\sum_{i=1}^m \epsilon_{i,k+1}^T(v_{i,k} - x^*) + \alpha_{k+1}^2\sum_{i=1}^m (C_i + \|\epsilon_{i,k+1}\|)^2.$$

Since X is bounded, by using $\|v_{i,k} - x^*\| \le \max_{x,y \in X}\|x - y\|$, taking the expectation and using the error bounds $\mathsf{E}\left[\|\epsilon_{i,k+1}\|^2\right] \le \bar{\nu}_i^2$ we obtain

$$\sum_{i=1}^m \mathsf{E}\left[\|v_{i,k+1} - x^*\|^2\right] \le \sum_{i=1}^m \mathsf{E}\left[\|v_{i,k} - x^*\|^2\right] - 2\alpha_{k+1}\left(\mathsf{E}[f(y_k)] - f^*\right)$$

$$+2\alpha_{k+1}\left(\max_{i \in V} C_i\right)\sum_{j=1}^m \mathsf{E}[\|y_k - w_{j,k}\|]$$

$$+2\alpha_{k+1}\max_{x,y \in X}\|x - y\|\sum_{i=1}^m \|\mathsf{E}[\epsilon_{i,k+1}]\| + \alpha_{k+1}^2\sum_{i=1}^m (C_i + \bar{\nu}_i)^2. \quad (5.3)$$

By rearranging the terms and summing over $k = 1, \dots, K$, for an arbitrary K, we obtain

$$2\sum_{k=1}^K \alpha_{k+1}\left((\mathsf{E}[f(y_k)] - f^*) - \left(\max_{i \in V} C_i\right)\sum_{j=1}^m \mathsf{E}[\|y_k - w_{j,k}\|]\right.$$

$$\left. - \max_{x,y \in X}\|x - y\|\sum_{i=1}^m \|\mathsf{E}[\epsilon_{i,k+1}]\| - \frac{m\alpha_{k+1}}{2}\left(\max_{i \in V}\{C_i + \bar{\nu}_i\}\right)^2\right)$$

$$\le \sum_{i=1}^m \mathsf{E}\left[\|v_{i,1} - x^*\|^2\right] - \sum_{i=1}^m \mathsf{E}\left[\|v_{i,K+1} - x^*\|^2\right] \le m\max_{x,y \in X}\|x - y\|^2.$$

Note that when $\alpha_{k+1} \to \alpha$ and $\alpha > 0$, we have $\sum_k \alpha_k = \infty$. When $\alpha = 0$, we have assumed that $\sum_k \alpha_k = \infty$. Therefore, by letting $K \to \infty$, we have

$$\liminf_{k \to \infty}\left(\mathsf{E}[f(y_k)] - \left(\max_{i \in V} C_i\right)\sum_{j=1}^m \mathsf{E}[\|y_k - w_{j,k}\|]\right.$$

$$\left. - \max_{x,y \in X}\|x - y\|\sum_{i=1}^m \|\mathsf{E}[\epsilon_{i,k+1}]\| - \frac{m\alpha_{k+1}}{2}\left(\max_{i \in V}\{C_i + \bar{\nu}_i\}\right)^2\right) \le f^*.$$

Using $\limsup_{k \to \infty}\|\mathsf{E}[\epsilon_{i,k+1}]\| = \bar{\mu}_i$ [see Eq. (5.2)] and $\lim_{k \to \infty}\alpha_k = \alpha$, we obtain

$$\liminf_{k \to \infty}\mathsf{E}[f(y_k)] \le f^* + \frac{m\alpha}{2}\left(\max_{i \in V}\{C_i + \bar{\nu}_i\}\right)^2 + \left(\max_{i \in V} C_i\right)\sum_{j=1}^m \limsup_{k \to \infty}\mathsf{E}[\|y_k - w_{j,k}\|]$$

$$+ \max_{x,y \in X}\|x - y\|\sum_{i=1}^m \bar{\mu}_i.$$

Next from the convexity inequality in (2.2) and the boundedness of the subgradients it follows that for all k and $j \in V$,

$$\mathsf{E}[f(w_{j,k}) - f(y_k)] \leq \left(\sum_{i=1}^{m} C_i \right) \mathsf{E}[\|y_k - w_{j,k}\|],$$

implying

$$\liminf_{k \to \infty} \mathsf{E}[f(w_{j,k})] \leq f^* + \frac{m\alpha}{2} \left(\max_{i \in V} \{C_i + \bar{\nu}_i\} \right)^2 + \left(\max_{i \in V} C_i \right) \sum_{j=1}^{m} \limsup_{k \to \infty} \mathsf{E}[\|y_k - w_{j,k}\|]$$

$$+ \left(\sum_{i=1}^{m} C_i \right) \limsup_{k \to \infty} \mathsf{E}[\|y_k - w_{j,k}\|] + \max_{x,y \in X} \|x - y\| \sum_{i=1}^{m} \bar{\mu}_i.$$

By Theorem 5.1, we have for all $j \in V$,

$$\limsup_{k \to \infty} \mathsf{E}[\|y_k - w_{j,k}\|] \leq \alpha \max_{i \in V} \{C_i + \bar{\nu}_i\} \left(2 + \frac{m\theta\beta}{1 - \beta} \right).$$

By using the preceding relation, we see that

$$\liminf_{k \to \infty} \mathsf{E}[f(w_{j,k})] \leq f^* + \frac{m\alpha}{2} \left(\max_{i \in V} \{C_i + \bar{\nu}_i\} \right)^2 + \max_{x,y \in X} \|x - y\| \sum_{i=1}^{m} \bar{\mu}_i$$

$$+ m\alpha \left(\max_{i \in V} C_i \right) \max_{i \in V} \{C_i + \bar{\nu}_i\} \left(2 + \frac{m\theta\beta}{1 - \beta} \right)$$

$$+ \alpha \left(\sum_{i=1}^{m} C_i \right) \max_{j \in V} \{C_j + \bar{\nu}_j\} \left(2 + \frac{m\theta\beta}{1 - \beta} \right)$$

$$\leq f^* + \max_{x,y \in X} \|x - y\| \sum_{i=1}^{m} \bar{\mu}_i + m\alpha \left(\max_{i \in V} \{C_i + \bar{\nu}_i\} \right)^2 \left(\frac{9}{2} + \frac{2m\theta\beta}{1 - \beta} \right).$$

□

   The network topology influences the error only through the term $\frac{\theta\beta}{1-\beta}$ and can hence be used as a figure of merit for comparing different topologies. For a network that is strongly connected at every time, [i.e., $Q = 1$ in Assumption 2] and when η in Assumption 3 does not depend on the number m of agents, the term $\frac{\theta\beta}{1-\beta}$ is of the order m^2 and the error bound scales as m^4.

   We next show that stronger bounds can be obtained for a specific weighted time averages of the iterates $w_{i,k}$. In particular, we investigate the limiting behavior of $\{f(z_{i,t})\}$, where $z_{i,t} = \frac{\sum_{k=1}^{t} \alpha_{k+1} w_{i,k}}{\sum_{k=1}^{t} \alpha_{k+1}}$. Note that agent i can locally and recursively evaluate $z_{i,t+1}$ from $z_{i,t}$ and $w_{i,t+1}$.

   THEOREM 5.4. *Consider the weighted time averages* $z_{j,t} = \frac{\sum_{k=1}^{t} \alpha_{k+1} w_{j,k}}{\sum_{k=1}^{t} \alpha_{k+1}}$ *for $j \in V$ and $t \geq 1$. Let the conditions of Theorem 5.3 hold. Then, we have for all $j \in V$,*

$$\limsup_{t \to \infty} \mathsf{E}[f(z_{j,t})] \leq f^* + \max_{x,y \in X} \|x - y\| \sum_{i=1}^{m} \bar{\mu}_i + m\alpha \left(\max_{i \in V} \{C_i + \bar{\nu}_i\} \right)^2 \left(\frac{9}{2} + \frac{2m\theta\beta}{1 - \beta} \right).$$

Proof. The relation in (5.3) of Theorem 5.3 is valid, and we have for any $x^* \in X^*$,

$$\sum_{i=1}^{m} \mathsf{E}\big[\|v_{i,k+1} - x^*\|^2\big] \le \sum_{i=1}^{m} \mathsf{E}\big[\|v_{i,k} - x^*\|^2\big] - 2\alpha_{k+1} \left(\mathsf{E}[f(y_k)] - f^*\right)$$

$$+ 2\alpha_{k+1} \left(\max_{i \in V} C_i\right) \sum_{\ell=1}^{m} \mathsf{E}[\|y_k - w_{\ell,k}\|]$$

$$+ 2\alpha_{k+1} \max_{x,y \in X} \|x - y\| \sum_{i=1}^{m} \|\mathsf{E}[\epsilon_{i,k+1}]\| + \alpha_{k+1}^2 \sum_{i=1}^{m} (C_i + \bar{\nu}_i)^2.$$

From the subgradient boundedness and the subgradient inequality in (2.2) we have for any j,

$$\mathsf{E}[f(y_k)] - \mathsf{E}[f(w_{j,k})] \ge - \left(\sum_{i=1}^{m} C_i\right) \mathsf{E}[\|y_k - w_{j,k}\|] \ge -m \left(\max_{i \in V} C_i\right) \mathsf{E}[\|y_k - w_{j,k}\|].$$

Therefore, we obtain

$$\sum_{i=1}^{m} \mathsf{E}\big[\|v_{i,k+1} - x^*\|^2\big] \le \sum_{i=1}^{m} \mathsf{E}\big[\|v_{i,k} - x^*\|^2\big] - 2\alpha_{k+1} \left(\mathsf{E}[f(w_{j,k})] - f^*\right)$$

$$+ 2\alpha_{k+1} \left(\max_{i \in V} C_i\right) \left(m\mathsf{E}[\|y_k - w_{j,k}\|] + \sum_{i=1}^{m} \mathsf{E}[\|y_k - w_{i,k}\|]\right)$$

$$+ 2\alpha_{k+1} \max_{x,y \in X} \|x - y\| \sum_{i=1}^{m} \|\mathsf{E}[\epsilon_{i,k+1}]\| + \alpha_{k+1}^2 \sum_{i=1}^{m} (C_i + \bar{\nu}_i)^2.$$

By re-arranging these terms, summing over $k = 1, \ldots, t$ and dividing with $2\sum_{k=1}^{t} \alpha_{k+1}$, we further obtain

$$\sum_{k=1}^{t} \frac{\alpha_{k+1}\mathsf{E}[f(w_{j,k})]}{\sum_{k=1}^{t} \alpha_{k+1}} \le f^* + \frac{1}{2\sum_{k=1}^{t} \alpha_{k+1}} \sum_{i=1}^{m} \mathsf{E}\big[\|v_{i,1} - x^*\|^2\big]$$

$$+ \sum_{k=1}^{t} \frac{\alpha_{k+1}\left(\max_{i \in V} C_i\right)\left(m\mathsf{E}[\|y_k - w_{j,k}\|] + \sum_{i=1}^{m} \mathsf{E}[\|y_k - w_{i,k}\|]\right)}{\sum_{k=1}^{t} \alpha_{k+1}}$$

$$+ \max_{x,y \in X} \|x - y\| \sum_{i=1}^{m} \frac{\sum_{k=1}^{t} \alpha_{k+1}\|\mathsf{E}[\epsilon_{i,k+1}]\|}{\sum_{k=1}^{t} \alpha_{k+1}} + \frac{\sum_{k=1}^{t} \alpha_{k+1}^2}{2\sum_{k=1}^{t} \alpha_{k+1}} \sum_{i=1}^{m} (C_i + \bar{\nu}_i)^2.$$

Next by the convexity of f note that

$$f(z_{j,t}) = f\left(\sum_{k=1}^{t} \frac{\alpha_{k+1} w_{j,k}}{\sum_{k=1}^{t} \alpha_{k+1}}\right) \le \sum_{k=1}^{t} \frac{\alpha_{k+1} f(w_{j,k})}{\sum_{k=1}^{t} \alpha_{k+1}}.$$

From the preceding two relations we obtain

$$\mathsf{E}[f(z_{j,t})] \le f^* + \frac{1}{2\sum_{k=1}^{t} \alpha_{k+1}} \sum_{i=1}^{m} \mathsf{E}\big[\|v_{i,1} - x^*\|^2\big]$$

$$+ \sum_{k=1}^{t} \frac{\alpha_{k+1}\left(\max_{i \in V} C_i\right)\left(m\mathsf{E}[\|y_k - w_{j,k}\|] + \sum_{i=1}^{m} \mathsf{E}[\|y_k - w_{i,k}\|]\right)}{\sum_{k=1}^{t} \alpha_{k+1}}$$

$$+ \max_{x,y \in X} \|x - y\| \sum_{i=1}^{m} \frac{\sum_{k=1}^{t} \alpha_{k+1}\|\mathsf{E}[\epsilon_{i,k+1}]\|}{\sum_{k=1}^{t} \alpha_{k+1}} + \frac{\sum_{k=1}^{t} \alpha_{k+1}^2}{2\sum_{k=1}^{t} \alpha_{k+1}} \sum_{i=1}^{m} (C_i + \bar{\nu}_i)^2. \quad (5.4)$$

First note that in the limit as $t \to \infty$, the second term in (5.4) converges to 0 since $\sum_{k=1}^{t} \alpha_{k+1} = \infty$. By using the results of Lemma 3.1c, for the remaining terms, we obtain

$$\limsup_{t\to\infty} \mathsf{E}[f(z_{j,t})] \leq f^* + \left(\max_{i\in V} C_i\right) \limsup_{k\to\infty} \left(m\mathsf{E}[\|y_k - w_{j,k}\|] + \sum_{i=1}^{m} \mathsf{E}[\|y_k - w_{i,k}\|]\right)$$
$$+ \max_{x,y\in X} \|x - y\| \sum_{i=1}^{m} \limsup_{k\to\infty} \|\mathsf{E}[\epsilon_{i,k+1}]\| + \frac{\alpha}{2} \sum_{i=1}^{m} (C_i + \bar{\nu}_i)^2.$$

By Theorem 5.1, we have for all $j \in V$,

$$\limsup_{k\to\infty} \mathsf{E}[\|y_k - w_{j,k}\|] \leq \alpha \max_{i\in V}\{C_i + \bar{\nu}_i\} \left(2 + \frac{m\theta\beta}{1-\beta}\right),$$

which when substituted in the preceding relation, yields

$$\limsup_{t\to\infty} \mathsf{E}[f(z_{j,t})] \leq f^* + 2m\alpha \left(\max_{i\in V} C_i\right) \max_{i\in V}\{C_i + \bar{\nu}_i\} \left(2 + \frac{m\theta\beta}{1-\beta}\right)$$
$$+ \max_{x,y\in X} \|x - y\| \sum_{i=1}^{m} \limsup_{k\to\infty} \|\mathsf{E}[\epsilon_{i,k+1}]\| + \frac{\alpha}{2} \sum_{i=1}^{m} (C_i + \bar{\nu}_i)^2$$
$$\leq f^* + \max_{x,y\in X} \|x - y\| \sum_{i=1}^{m} \limsup_{k\to\infty} \|\mathsf{E}[\epsilon_{i,k+1}]\|$$
$$+ m\alpha \left(\max_{i\in V}\{C_i + \bar{\nu}_i\}\right)^2 \left(\frac{9}{2} + \frac{2m\theta\beta}{1-\beta}\right).$$

□

The error bounds in Theorems 5.3 and 5.4 have the same form, but they apply to different sequences of function evaluations. Furthermore, in Theorem 5.4, the bound is for *all* subsequences of $\mathsf{E}[f(z_{i,k})]$ for each agent i. In contrast, in Theorem 5.3, the bound is only for *a* subsequence of $\mathsf{E}[f(z_{i,k})]$ for each agent i. Theorem 5.4 demonstrates that, due to the convexity of the objective function f, there is an advantage when agents are using the running averages of their iterates.

When the error[5] moments $\|\mathsf{E}[\epsilon_{i,k+1}]\|$ converge to zero as $k \to \infty$, and the stepsize converges to zero $[\alpha = 0]$, Theorems 5.3 and 5.4 yield respectively

$$\liminf_{k\to\infty} \mathsf{E}[f(w_{j,k})] = f^* \qquad \text{and} \qquad \lim_{k\to\infty} \mathsf{E}[f(z_{j,k})] = f^*.$$

When a constant stepsize α is used, the vector $z_{j,t}$ is simply the running average of all the iterates of agent j until time t, i.e., $z_{j,t} = \frac{1}{t}\sum_{k=1}^{t} w_{j,k}$. For this case, with

---

[5] When the moments $\|\mathsf{E}[\epsilon_{i,k+1}]\|$ are zero, it can be seen that the results of Theorems 5.3 and 5.4 hold when the boundedness of X is replaced by the weaker assumption that the subgradients of each f_i are bounded over X.

zero mean errors, the relation in (5.4) reduces to

$$
\mathsf{E}[f(z_{j,t})] \leq f^* + \frac{1}{2t\alpha} \sum_{i=1}^{m} \mathsf{E}\big[\|v_{i,1} - x^*\|^2\big]
$$
$$
+ \left(\max_{i \in V} C_i \right) \frac{1}{t} \sum_{k=1}^{t} \left(m\mathsf{E}[\|y_k - w_{j,k}\|] + \sum_{i=1}^{m} \mathsf{E}[\|y_k - w_{i,k}\|] \right)
$$
$$
+ \frac{\alpha}{2} \sum_{i=1}^{m} (C_i + \bar{\nu}_i)^2. \tag{5.5}
$$

This can be used to derive an estimate per iteration, as seen in the following.

COROLLARY 5.5. *Under the conditions of Theorem 5.3 with* $\|\mathsf{E}[\epsilon_{i,k+1}]\| = 0$ *and* $\alpha_k = 0$ *for all i and k, for the average sequences* $\{z_{j,k}\}$ *we have for all t and j,*

$$
\mathsf{E}[f(z_{j,t})] \leq f^* + \frac{1}{2t\alpha} \sum_{i=1}^{m} \mathsf{E}\big[\|v_{i,1} - x^*\|^2\big] + \frac{2m^2\theta\beta^2}{t(1-\beta)} \left(\max_{i \in V} C_i \right) \left(\max_{i \in V} \|w_{i,0}\| \right)
$$
$$
+ m\alpha \left(\max_{i \in V}\{C_i + \bar{\nu}_i\} \right)^2 \left(\frac{9}{2} + \frac{2m\theta\beta}{1-\beta} \right).
$$

Proof. Taking the expectation in the relation of Lemma 4.1, we obtain

$$
\mathsf{E}[\|y_{k+1} - w_{j,k+1}\|] \leq m\theta\beta^{k+1} \max_{i \in V} \|w_{i,0}\| + m\alpha\theta\beta \left(\max_{i \in V}\{C_i + \bar{\nu}_i\} \right) \sum_{\ell=1}^{k} \beta^{k-\ell}
$$
$$
+ 2\alpha \max_{i \in V}\{C_i + \bar{\nu}_i\}
$$
$$
\leq m\theta\beta^{k+1} \max_{i \in V} \|w_{i,0}\| + \alpha \left(\max_{i \in V}\{C_i + \bar{\nu}_i\} \right) \left(2 + \frac{m\theta\beta}{1-\beta} \right).
$$

Combining the preceding relation with the inequality in (5.5), and using $\sum_{k=1}^{t} \beta^{k+1} \leq \frac{\beta^2}{1-\beta}$, we obtain

$$
\mathsf{E}[f(z_{j,t})] \leq f^* + \frac{1}{2t\alpha} \sum_{i=1}^{m} \mathsf{E}\big[\|v_{i,1} - x^*\|^2\big] + \frac{2m^2\theta\beta^2}{t(1-\beta)} \left(\max_{i \in V} C_i \right) \left(\max_{i \in V} \|w_{i,0}\| \right)
$$
$$
+ m\alpha \left(\max_{i \in V}\{C_i + \bar{\nu}_i\} \right)^2 \left(\frac{9}{2} + \frac{2m\theta\beta}{1-\beta} \right).
$$

□

The preceding equation provides a bound on the algorithm's performance at each iteration. The bound can be used in obtaining stopping rules for the algorithm. For example, consider the error free case ($\bar{\nu}_i = 0$) and suppose that the goal is to determine the number of iterations required for agents to find a point in the ϵ-optimal set, i.e., in the set $X_\epsilon = \{x \in X : f(x) \leq f^* + \epsilon\}$. Minimizing the bound in Corollary 5.5 over different stepsize values α, we can show that ϵ-optimality can be achieved in $N_\epsilon = \left\lceil \frac{1}{\psi_\epsilon^2} \right\rceil$ iterations with a stepsize $\alpha_\epsilon = \frac{\sqrt{A}\psi_\epsilon}{\sqrt{C}}$, where ψ_ϵ is the positive root of the quadratic equation

$$
Bx^2 + 2\sqrt{AC}x - \epsilon = 0,
$$

and A, B and C are

$$A = \frac{1}{2} \sum_{i=1}^{m} \|v_{i,1} - x^*\|^2, \qquad B = \frac{2m^2\theta\beta^2}{1-\beta} \left(\max_{i \in V} C_i\right) \left(\max_{i \in V} \|w_{i,0}\|\right),$$

$$C = m \left(\max_{i \in V} \{C_i + \bar{\nu}_i\}\right)^2 \left(\frac{9}{2} + \frac{2m\theta\beta}{1-\beta}\right).$$

Since ψ_ϵ scales as $\sqrt{\epsilon}$, we can conclude that N_ϵ scales as $\frac{1}{\epsilon^2}$. Equivalently, we can say that the level ϵ of sub-optimality diminishes inversely with the square root of the number of iterations.

6. Almost sure and mean square convergence. In this section, we impose some additional assumptions on the subgradient errors to obtain almost sure consensus among the agents and almost sure convergence of the iterates to an optimal solution of (2.1). Towards this, define F_k to be the σ-algebra $\sigma\left(\epsilon_{i,\ell}; \ i \in V, 0 \leq \ell \leq k\right)$ generated by the errors in the agent system up to time k. In other words, F_k captures the history of the errors until the end of time k. We use the following assumption on the subgradient errors $\epsilon_{i,k}$.

ASSUMPTION 5. *There are scalars ν_i such that $\mathsf{E}\left[\|\epsilon_{i,k+1}\|^2 \mid F_k\right] \leq \nu_i^2$ for all k with probability 1.*

Note that Assumption 5 is stronger than Assumption 4. Furthermore, when the errors are independent across iterations and across agents, Assumption 5 reduces to Assumption 4.

We start by analyzing the agents' disagreements measured in terms of distances $\|y_k - w_{j,k}\|$. We have the following result.

THEOREM 6.1. *Let Assumptions 1a, 2, 3 and 5 hold. Suppose that the subgradients of each f_i are uniformly bounded over X, i.e., for each $i \in V$ there is C_i such that*

$$\|\nabla f_i(x)\| \leq C_i \qquad \text{for all } x \in X.$$

If $\sum_{k=0}^{\infty} \alpha_{k+1}^2 < \infty$, then with probability 1,

$$\sum_{k=1}^{\infty} \alpha_{k+2} \|y_{k+1} - w_{j,k+1}\| < \infty \qquad \text{for all } j \in V.$$

Furthermore, for all $j \in V$, we have $\lim_{k \to \infty} \|y_{k+1} - w_{j,k+1}\| = 0$ with probability 1 and in mean square.

Proof. By Lemma 4.1 and the subgradient boundedness, we have for all $j \in V$,

$$\|y_{k+1} - w_{j,k+1}\| \leq m\theta\beta^{k+1} \max_{i \in V} \|w_{i,0}\| + \theta \sum_{\ell=1}^{k} \beta^{k+1-\ell} \sum_{i=1}^{m} \alpha_\ell \left(C_i + \|\epsilon_{i,\ell}\|\right)$$

$$+ \frac{1}{m} \sum_{i=1}^{m} \alpha_{k+1} \left(C_i + \|\epsilon_{i,k+1}\|\right) + \alpha_{k+1} \left(C_j + \|\epsilon_{j,k+1}\|\right).$$

Using the inequalities

$$\alpha_{k+2}\alpha_\ell \left(C_i + \|\epsilon_{i,\ell}\|\right) \leq \frac{1}{2} \left(\alpha_{k+2}^2 + \alpha_\ell^2 \left(C_i + \|\epsilon_{i,\ell}\|\right)^2\right)$$

and $(C_i + \|\epsilon_{i,\ell}\|)^2 \leq 2C_i^2 + 2\|\epsilon_{i,\ell}\|^2$, we obtain

$$\alpha_{k+2}\|y_{k+1} - w_{j,k+1}\| \leq \alpha_{k+2}\, m\theta\beta^{k+1} \max_{i\in V} \|w_{i,0}\|$$

$$+ \theta \sum_{\ell=1}^{k} \beta^{k+1-\ell} \sum_{i=1}^{m} \left(\frac{1}{2}\alpha_{k+2}^2 + \alpha_\ell^2 \left(C_i^2 + \|\epsilon_{i,\ell}\|^2 \right) \right)$$

$$+ \frac{1}{m} \sum_{i=1}^{m} \left(\frac{1}{2}\alpha_{k+2}^2 + \alpha_{k+1}^2 \left(C_i^2 + \|\epsilon_{i,k+1}\|^2 \right) \right)$$

$$+ \frac{1}{2}\alpha_{k+2}^2 + \alpha_{k+1}^2 \left(C_j^2 + \|\epsilon_{j,k+1}\|^2 \right).$$

By using the inequalities $\sum_{\ell=1}^{k} \beta^{k+1-\ell} \leq \frac{\beta}{1-\beta}$ for all $k \geq 1$ and $\frac{1}{2m} + \frac{1}{2} \leq 1$, and by grouping the terms accordingly, from the preceding relation we have

$$\alpha_{k+2}\|y_{k+1} - w_{j,k+1}\| \leq \alpha_{k+2}\, m\theta\beta^{k+1} \max_{i\in V} \|w_{i,0}\| + \left(1 + \frac{m\theta\beta}{2(1-\beta)} \right) \alpha_{k+2}^2$$

$$+ \theta \sum_{\ell=1}^{k} \alpha_\ell^2 \beta^{k+1-\ell} \sum_{i=1}^{m} \left(C_i^2 + \|\epsilon_{i,\ell}\|^2 \right)$$

$$+ \frac{1}{m}\alpha_{k+1}^2 \sum_{i=1}^{m} \left(C_i^2 + \|\epsilon_{i,k+1}\|^2 \right) + \alpha_{k+1}^2 \left(C_j^2 + \|\epsilon_{j,k+1}\|^2 \right).$$

Taking the conditional expectation and using $\mathsf{E}\big[\|\epsilon_{i,\ell}\|^2 \mid F_{\ell-1}\big] \leq \nu_i^2$, and then taking the expectation again, we obtain

$$\mathsf{E}[\alpha_{k+2}\|y_{k+1} - w_{j,k+1}\|] \leq \alpha_{k+2}\, m\theta\beta^{k+1} \max_{i\in V} \|w_{i,0}\| + \left(1 + \frac{m\theta\beta}{2(1-\beta)} \right) \alpha_{k+2}^2$$

$$+ \theta \left(\sum_{i=1}^{m} \left(C_i^2 + \nu_i^2 \right) \right) \sum_{\ell=1}^{k} \alpha_\ell^2 \beta^{k+1-\ell}$$

$$+ \frac{1}{m}\alpha_{k+1}^2 \sum_{i=1}^{m} \left(C_i^2 + \nu_i^2 \right) + \alpha_{k+1}^2 \left(C_j^2 + \nu_j^2 \right).$$

Since $\sum_k \alpha_k^2 < \infty$ (and hence $\{\alpha_k\}$ bounded), the first two terms and the last two terms are summable. Furthermore, in view of Lemma 3.1 [part (b)], we have

$$\sum_{k=1}^{\infty}\sum_{\ell=1}^{k} \beta^{k+1-\ell}\alpha_\ell^2 < \infty.$$

Thus, the third term is also summable. Hence $\sum_{k=1}^{\infty} \mathsf{E}[\alpha_{k+2}\|y_{k+1} - w_{j,k+1}\|] < \infty$. From the monotone convergence theorem [4], it follows that

$$\mathsf{E}\left[\sum_{k=1}^{\infty} \alpha_{k+2}\|y_{k+1} - w_{j,k+1}\| \right] = \sum_{k=1}^{\infty} \mathsf{E}[\alpha_{k+2}\|y_{k+1} - w_{j,k+1}\|],$$

and it is hence finite for all j. If the expected value of a random variable is finite, then the variable has to be finite with probability 1; thus, with probability 1,

$$\sum_{k=1}^{\infty} \alpha_{k+2}\|y_{k+1} - w_{j,k+1}\| < \infty \qquad \text{for all } j \in V. \tag{6.1}$$

We now show that $\lim_{k\to\infty} \|y_k - w_{j,k}\| = 0$ with probability 1 for all $j \in V$. Note that the conditions of Theorem 5.1 are satisfied with $\bar{\nu}_i = \nu_i$ and $\alpha = 0$. Therefore, $\|y_k - w_{j,k}\|$ converges to 0 in the mean and from (Fatou's) Lemma 3.3 it follows that

$$0 \leq \mathsf{E}\left[\liminf_{k\to\infty} \|y_k - w_{j,k}\|\right] \leq \liminf_{k\to\infty} \mathsf{E}[\|y_k - w_{j,k}\|] = 0,$$

and hence $\mathsf{E}[\liminf_{k\to\infty} \|y_k - w_{j,k}\|] = 0$. Therefore, with probability 1,

$$\liminf_{k\to\infty} \|y_k - w_{j,k}\| = 0. \tag{6.2}$$

To complete the proof, in view of (6.2) it suffices to show that $\|y_k - w_{j,k}\|$ converges with probability 1. To show this, we define

$$r_{i,k+1} = \sum_{j=1}^{m} a_{i,j}(k+1)w_j(k) - \alpha_{k+1}\left(\nabla f_i\left(v_{i,k}\right) + \epsilon_{i,k+1}\right),$$

and note that $P_X[r_{i,k+1}] = w_{i,k+1}$ [see (2.3) and (2.4)]. Since $y_k = \frac{1}{m}\sum_{i=1}^{m} w_{i,k}$ and the set X is convex, it follows that $y_k \in X$ for all k. Therefore, by the non-expansive property of the Euclidean projection in (3.4), we have $\|w_{i,k+1} - y_k\|^2 \leq \|r_{i,k+1} - y_k\|^2$ for all $i \in V$ and all k. Summing these relations over all i, we obtain

$$\sum_{i=1}^{m} \|w_{i,k+1} - y_k\|^2 \leq \sum_{i=1}^{m} \|r_{i,k+1} - y_k\|^2 \qquad \text{for all } k.$$

From $y_{k+1} = \frac{1}{m}\sum_{i=1}^{m} w_{i,k+1}$ and the fact that the average of vectors minimizes the sum of distances between each vector and arbitrary vector in $\Re^n$ [cf. Eq (3.1)], we further obtain

$$\sum_{i=1}^{m} \|w_{i,k+1} - y_{k+1}\|^2 \leq \sum_{i=1}^{m} \|w_{i,k+1} - y_k\|^2.$$

Therefore, for all k,

$$\sum_{i=1}^{m} \|w_{i,k+1} - y_{k+1}\|^2 \leq \sum_{i=1}^{m} \|r_{i,k+1} - y_k\|^2. \tag{6.3}$$

We next relate $\sum_{i=1}^{m} \|r_{i,k+1} - y_k\|^2$ to $\sum_{i=1}^{m} \|w_{i,k} - y_k\|^2$. From the definition of $r_{i,k+1}$ and the equality $\sum_{j=1}^{m} a_{i,j}(k+1) = 1$ [cf. Assumption 3b], we have

$$r_{i,k+1} - y_k = \sum_{j=1}^{m} a_{i,j}(k+1)\left(w_{j,k} - y_k\right) - \alpha_{k+1}\left(\nabla f_i(v_{i,k}) + \epsilon_{i,k+1}\right)$$

By Assumption 3a and 3b, we have that the weights $a_{i,j}(k+1), j \in V$ yield a convex combination. Thus, by the convexity of the norm [(3.2) and (3.3)] and by the

subgradient boundedness, we have

$$\|r_{i,k+1} - y_k\|^2 \le \sum_{j=1}^m a_{i,j}(k+1)\,\|w_{j,k} - y_k\|^2 + \alpha_{k+1}^2\,\|\nabla f_i(v_{i,k}) + \epsilon_{i,k+1}\|^2$$

$$+ 2\alpha_{k+1}\,\|\nabla f_i(v_{i,k}) + \epsilon_{i,k+1}\|\sum_{j=1}^m a_{i,j}(k+1)\,\|w_{j,k} - y_k\|$$

$$\le \sum_{j=1}^m a_{i,j}(k+1)\,\|w_{j,k} - y_k\|^2 + 2\alpha_{k+1}^2\left(C_i^2 + \|\epsilon_{i,k+1}\|^2\right)$$

$$+ 2\alpha_{k+1}\left(C_i + \|\epsilon_{i,k+1}\|\right)\sum_{j=1}^m a_{i,j}(k+1)\,\|w_{j,k} - y_k\|.$$

Summing over all i and using $\sum_{i=1}^m a_{i,j}(k+1) = 1$ [cf. Assumption 3d], we obtain

$$\sum_{i=1}^m \|r_{i,k+1} - y_k\|^2 \le \sum_{j=1}^m \|w_{j,k} - y_k\|^2 + 2\alpha_{k+1}^2 \sum_{i=1}^m \left(C_i^2 + \|\epsilon_{i,k+1}\|^2\right)$$

$$+ 2\alpha_{k+1} \sum_{i=1}^m \left(C_i + \|\epsilon_{i,k+1}\|\right)\sum_{j=1}^m a_{i,j}(k+1)\,\|w_{j,k} - y_k\|.$$

Using this in (6.3) and taking the conditional expectation, we see that for all k, we have with probability 1,

$$\sum_{i=1}^m \mathsf{E}\big[\|w_{i,k+1} - y_{k+1}\|^2 \mid F_k\big] \le \sum_{i=1}^m \|w_{i,k} - y_k\|^2 + 2\alpha_{k+1}^2 \sum_{i=1}^m \left(C_i^2 + \nu_i^2\right)$$

$$+ 2\alpha_{k+1} \sum_{i=1}^m \left(C_i + \nu_i\right)\sum_{j=1}^m \|w_{j,k} - y_k\|, \qquad (6.4)$$

where we use $a_{i,j}(k+1) \le 1$ for all i, j and k, and the relations $\mathsf{E}\big[\|\epsilon_{i,k+1}\|^2 \mid F_k\big] \le \nu_i^2$, $\mathsf{E}[\|\epsilon_{i,k+1}\| \mid F_k] \le \nu_i$ holding with probability 1.

We now apply Theorem 3.4 to the relation in (6.4). To verify that the conditions of Theorem 3.4 are satisfied, note that the stepsize satisfies $\sum_{k=1}^\infty \alpha_{k+1}^2 < \infty$ for all $i \in V$. We also have $\sum_{k=1}^\infty \alpha_{k+1}\,\|w_{j,k} - y_k\| < \infty$ with probability 1 [cf. (6.1)]. Therefore, the relation in (6.4) satisfies the conditions of Theorem 3.4 with $\zeta_k = D_k = 0$, thus implying that $\|w_{j,k} - y_k\|$ converges with probability 1 for every $j \in V$. □

Let us compare Theorem 6.1 and Corollary 5.2. Corollary 5.2 provided sufficient conditions for the different agents to have consensus in the mean. Theorem 6.1 strengthens this to consensus with probability 1 and in mean square sense, for a smaller class of stepsize sequences under a stricter assumption.

We next show that the consensus vector is actually in the optimal set, provided that the optimal set is nonempty and the conditional expectations $\|\mathsf{E}[\epsilon_{i,k+1} \mid F_k]\|$ are diminishing.

THEOREM 6.2. *Let Assumptions 1, 2, 3 and 5 hold. Suppose that the subgradients of each f_i are uniformly bounded over X, i.e., for each $i \in V$ there is C_i such that*

$$\|\nabla f_i(x)\| \le C_i \qquad \text{for all } x \in X.$$

Also, assume that $\sum_{k=0}^\infty \|\mathsf{E}[\epsilon_{i,k+1} \mid F_k]\|^2 < \infty$ for all $i \in V$. Further, let the stepsize sequence $\{\alpha_k\}$ be such that $\sum_{k=1}^\infty \alpha_k = \infty$ and $\sum_{k=1}^\infty \alpha_k^2 < \infty$. Then, if the optimal

set X^* is nonempty, the iterate sequence $\{w_{i,k}\}$ of each agent $i \in V$ converges to the same optimal point with probability 1 and in mean square.

   Proof. Observe that the conditions of Lemma 4.2 are satisfied. Letting $z = x^*$ for some $x^* \in X^*$, taking conditional expectations and using the bounds on the error moments, we obtain for any $x^* \in X^*$ and any k, with probability 1,

$$\sum_{i=1}^{m} \mathsf{E}\left[\|v_{i,k+1} - x^*\|^2 \mid F_k\right] \leq \sum_{i=1}^{m} \|v_{i,k} - x^*\|^2 - 2\alpha_{k+1}\left(f(y_k) - f^*\right)$$

$$+ 2\alpha_{k+1}\left(\max_{i \in V} C_i\right)\sum_{j=1}^{m} \|y_k - w_{j,k}\|$$

$$+ 2\alpha_{k+1}\sum_{i=1}^{m} \mu_{i,k+1}\|v_{i,k} - x^*\| + \alpha_{k+1}^2 \sum_{i=1}^{m} (C_i + \nu_i)^2,$$

where $f^* = f(x^*)$, and we use the notation $\mu_{i,k+1} = \|\mathsf{E}[\epsilon_{i,k+1} \mid F_k]\|$. Using the inequality

$$2\alpha_{k+1}\mu_{i,k+1}\|v_{i,k} - x^*\| \leq \alpha_{k+1}^2\|v_{i,k} - x^*\|^2 + \mu_{i,k+1}^2,$$

we obtain with probability 1,

$$\sum_{i=1}^{m} \mathsf{E}\left[\|v_{i,k+1} - x^*\|^2 \mid F_k\right] \leq \sum_{i=1}^{m} \left(1 + \alpha_{k+1}^2\right)\|v_{i,k} - x^*\|^2$$

$$- 2\alpha_{k+1}\left((f(y_k) - f^*) - \left(\max_{i \in V} C_i\right)\sum_{j=1}^{m} \|y_k - w_{j,k}\|\right.$$

$$+ \sum_{i=1}^{m} \mu_{i,k+1}^2 - \frac{1}{2}\alpha_{k+1}\sum_{i=1}^{m} (C_i + \nu_i)^2\bigg). \qquad (6.5)$$

By Theorem 6.1, we have with probability 1,

$$\sum_{k} \alpha_{k+1}\|w_{j,k} - y_k\| < \infty.$$

Further, since $\sum_k \mu_{i,k}^2 < \infty$ and $\sum_k \alpha_k^2 < \infty$ with probability 1, the relation in (6.5) satisfies the conditions of Theorem 3.4. We therefore have

$$\sum_{k} \alpha_k(f(y_k) - f^*) < \infty, \qquad (6.6)$$

and $\|v_{i,k} - x^*\|$ converges with probability 1 and in mean square. In addition, by Theorem 6.1, we have $\lim_{k \to \infty} \|w_{i,k} - y_k\| = 0$ for all i, with probability 1. Hence, $\lim_{k \to \infty} \|v_{i,k} - y_k\| \to 0$ for all i, with probability 1. Therefore, $\|y_k - x^*\|$ converges with probability 1 for any $x^* \in X^*$. Moreover, from (6.6) and the fact that $\sum_k \alpha_k = \infty$, by continuity of f, it follows that y_k, and hence $w_{i,k}$, must converge to a vector in X^* with probability 1 and in mean square. $\square$

   Note that the result of Theorem 6.2 holds without assuming compactness of the constraint set X. This was possible due to the assumption that both the stepsize α_k and the norms $\|\mathsf{E}[\epsilon_{i,k+1} \mid F_k]\|$ of the conditional errors are square summable. In addition, note that the result of Theorem 6.2 remains valid when the condition $\sum_{k=0}^{\infty} \|\mathsf{E}[\epsilon_{i,k+1} \mid F_k]\|^2 < \infty$ for all i is replaced with $\sum_{k=0}^{\infty} \alpha_{k+1}\|\mathsf{E}[\epsilon_{i,k+1} \mid F_k]\| < \infty$ for all i.

7. Implications. The primary source of stochastic errors in the subgradient evaluation is when the objective function is not completely known and has some randomness in it. Such settings arise in sensor network applications that involve distributed and recursive estimation [24].

Let the function $f_i(x)$ be given by $f_i(x) = \mathsf{E}[g_i(x, R_i)]$, where R_i is a random variable whose statistics are independent of x. The statistics of R_i are not available to agent i and hence the function f_i is not known to agent i. Instead, agent i observes samples of R_i in time. Thus, in a subgradient algorithm for minimizing the function, the subgradient must be suitably approximated using the observed samples. In the Robbins-Monro stochastic approximation [22], the subgradient $\nabla f_i(x)$ is approximated by $\nabla g_i(x, r_i)$, where r_i denotes a sample of R_i. The associated distributed Robbins-Monro stochastic optimization algorithm is

$$w_{i,k+1} = P_X \left[v_{i,k} - \alpha_{k+1} \nabla g_i \left(v_{i,k}, r_{i,k+1} \right) \right], \tag{7.1}$$

where $r_{i,k+1}$ is a sample of R_i obtained at time k. The expression for the error is

$$\epsilon_{i,k+1} = \nabla g_i(v_{i,k}, r_{i,k+1}) - \mathsf{E}[\nabla g_i(v_{i,k}, R_i)].$$

If the samples obtained across iterations are independent then

$$\mathsf{E}[\epsilon_{i,k+1} \mid F_k] = \mathsf{E}[\epsilon_{i,k+1} \mid v_{i,k}] = 0.$$

If in addition, $\mathsf{Var}[\nabla g_i(x, R_i)]$ is bounded for all $x \in X$ then the conditions of Theorems 5.3, 5.4 and 6.2 are satisfied.

Let us next consider the case when $f_i(x) = \mathsf{E}[g_i(x, R_i(x))]$, where $R_i(x)$ is a random variable that is parameterized by x. To keep the discussion simple, let us assume that $x \in \Re$. As in the preceding case, the statistics of $R_i(x)$ are not known to agent i, but the agent can obtain samples of $R_i(x)$ for any value of x. In the Kiefer-Wolfowitz approximation [22],

$$\nabla f_i(x) \approx \frac{g_i\left(x, r_i(x + \beta)\right) - g_i\left(x, r_i(x)\right)}{\beta},$$

where $r_i(x)$ is a sample of the random variable $R_i(x)$. The corresponding distributed optimization algorithm is

$$w_{i,k+1} = P_X \left[v_{i,k} - \alpha_{k+1} \frac{g_i\left(v_{i,k}, r_i(v_{i,k} + \beta_{i,k+1})\right) - g_i\left(v_{i,k}, r_i(v_{i,k})\right)}{\beta_{i,k+1}} \right],$$

where $\beta_{i,k+1}$ is a positive scalar. In this case, the error is

$$\epsilon_{i,k+1} = \frac{g_i\left(v_{i,k}, r_i(v_{i,k} + \beta_{i,k+1})\right) - g_i\left(v_{i,k}, r_i(v_{i,k})\right)}{\beta_{i,k+1}} - \nabla f_i(v_{i,k}).$$

If the function g_i is differentiable then $\mathsf{E}[\epsilon_{i,k+1} \mid v_{i,k}]$ is of the order $\beta_{i,k+1}$. Thus, the conditions on the mean value of the errors can be controlled through the sequence $\{\beta_{i,k}\}$ and the conditions in Theorems 5.3, 5.4 and 6.2 can be met by suitably choosing the sequence $\{\beta_{i,k}\}$.

8. Discussion. We studied the effects of stochastic subgradient errors on distributed algorithm for network of agents with time-varying connectivity. We first considered very general errors with bounded second moments and obtained explicit bounds on the agent disagreements and on the expected deviation of the limiting function value from the optimal. The bounds are explicitly given as a function of the network properties, objective function and the error moments. For networks that are connected at all times and η is independent of the size of the network, the bound scales as $\alpha \left(\max_{i \in V}\{C_i + \nu_i\}\right)^2 m^4$, where m is the number of agents in the network, α is the stepsize limit, and C_i and ν_i^2 are respectively the subgradient norm bound and the bound on the second moment of the subgradient errors for agent i. For the constant stepsize case, we obtained a bound on the performance of the algorithm after a finite number of iterations. There, we showed that deviation from the "error-bound" diminishes at rate $\frac{1}{t}$, where t is the number of iterations. Finally, we proved that when the expected error and the stepsize converge to 0 sufficiently fast, the agents reach a consensus and the iterate sequences of agents converge to a common optimal point with probability 1 and in mean square.

We make the following remarks. First, it can be shown that the disagreement results in Corollary 5.2 and Theorem 6.1 hold even when the agents use non-identical stepsizes. However, with non-identical agent stepsizes there is no guarantee that the sum of the objectives rather than a weighted sum, is minimized.

Future work includes several important extensions of the distributed model studied here. At first, we have assumed no communication delays between the agents and synchronous processing. An important extension is to consider the properties of the algorithm in asynchronous networks with communication delays, as in [30]. At second, we assumed perfect communication scenario, i.e., noiseless communication links. In wireless network applications, the links are typically noisy and this has to be taken into consideration. At third, we have considered the class of convex functions. This restricts the number of possible applications for the algorithm. Further research is to develop distributed algorithms when the functions f_i are not convex.

REFERENCES

[1] D. P. Bertsekas, A. Nedić, and A. Ozdalgar, *Convex Analysis and Optimization*, Athena Scientific, 2003.

[2] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and distributed computation: Numerical methods*, Athena Scientific, 1997.

[3] ———, *Gradient convergence in gradient methods with errors*, SIAM Journal of Optimization, 10 (2000), pp. 627–642.

[4] P. Billingsley, *Probability and Measure*, John Wiley and Sons, 1979.

[5] D. Blatt, A. O. Hero, and H. Gauchman, *A convergent incremental gradient method with constant stepsize*, SIAM Journal of Optimization, 18 (2007), pp. 29–51.

[6] Y. Ermoliev, *Stochastic programming methods*, Nauka, Moscow, 1976.

[7] ———, *Stochastic quasi-gradient methods and their application to system optimization*, Stochastics, 9 (1983), pp. 1–36.

[8] ———, *Stochastic quazigradient methods*, Numerical Techniques for Stochastic Optimization, (Eds.) Y. Ermoliev and R. J-B Wets, Springer-Verlag, N.Y., (Eds.) Y. Ermoliev and R. J-B Wets, Springer-Verlag, N.Y. (1988), pp. 141–186.

[9] A. A. Gaivoronski, *Convergence properties of backpropogation for neural nets via theory of stochastic gradient methods. Part 1.*, Optimization Methods and Software, 4 (1994), pp. 117–134.

[10] M. Huang and J. H. Manton, *Stochastic approximation for consensus seeking: Mean square and almost sure convergence*, in Proceedings of the 46th IEEE Conference on Decision and Control, 2007.

[11] A. Jadbabaie, J. Lin, and S. Morse, *Coordination of groups of mobile autonomous agents*

using nearest neighbor rules, IEEE Transactions on Automatic Control, 48 (2003), pp. 998–1001.

[12] B. JOHANSSON, M. RABI, AND M. JOHANSSON, *A simple peer-to-peer algorithm for distributed optimization in sensor networks*, in Proceedings of the 46th IEEE Conference on Decision and Control, 2007, pp. 4705–4710.

[13] S. KAR AND J. MOURA, *Distributed consensus algorithms in sensor networks: Link and channel noise.* http://arxiv.org/abs/0711.3915, 2007.

[14] K. C. KIWIEL, *Convergence of approximate and incremental subgradient methods for convex optimization*, SIAM Journal on Optimization, 14 (2003), pp. 807–840.

[15] I. LOBEL AND A. OZDAGLAR, *Distributed subgradient methods over random networks.* MIT, LIDS Technical Report 2800, submitted, 2008.

[16] A. NEDIĆ AND D. P. BERTSEKAS, *Incremental subgradient method for nondifferentiable optimization*, SIAM Journal of Optimization, 12 (2001), pp. 109–138.

[17] ———, *The effect of deterministic noise in sub-gradient methods*, tech. report, Lab. for Information and Decision Systems, MIT, 2007.

[18] A. NEDIĆ, A. OLSHEVSKY, A. OZDAGLAR, AND J. N. TSITSIKLIS, *Distributed subgradient algorithms and quantization effects.* http://arxiv.org/abs/0803.1202, 2008.

[19] A. NEDIĆ AND A. OZDAGLAR, *Distributed sub-gradient methods for multi-agent optimization*, To appear in Transactions on Automatic Control, (2008).

[20] A. NEDIĆ, A. OZDAGLAR, AND P. A. PARRILO, *Constrained consensus.* Lab. for Information and Decision Systems, MIT, Report 2779, 2008.

[21] A. OLSHEVSKY AND J. N. TSITSIKLIS, *Convergence speed in distributed consensus and control*, To appear in SIAM Journal on Control and Optimization, (2008).

[22] B. T. POLYAK, *Introduction to optimization*, Optimization Software Inc., 1987.

[23] M. G. RABBAT AND R. D. NOWAK, *Quantized incremental algorithms for distributed optimization*, IEEE Journal on Select Areas in Communications, 23 (2005), pp. 798–808.

[24] S. SUNDHAR RAM, A. NEDIĆ, AND V. V. VEERAVALLI, *Stochastic incremental gradient descent for estimation in sensor networks*, in Proceedings of the Asilomar Conference on Signals, Systems, and Computers, 2007, pp. 582–586.

[25] S. SUNDHAR RAM, A. NEDIĆ, AND V. V. VEERAVALLI, *Incremental stochastic sub-gradient algorithms for convex optimization.* Available at http://arxiv.org/abs/0806.1092, 2008.

[26] S. SUNDHAR RAM, V. VEERAVALLI, AND A. NEDIĆ, *Distributed and non-autonomous power control through distributed convex optimization.* submitted, 2008.

[27] R. T. ROCKAFELLAR, *Convex Analysis*, Princeton University Press, 1970.

[28] M. V. SOLODOV AND S. K. ZAVRIEV, *Error stability properties of generalized gradient-type algorithms*, Journal of Optimization Theory and Applications, 98 (1998), pp. 663–680.

[29] D. P. SPANOS, R. OLFATI-SABER, AND R. M. MURRAY, *Approximate distributed Kalman filtering in sensor networks with quantifiable performance*, in Proceedings of IEEE International Conference on Information Processing in Sensor Networks, 2005, pp. 133–139.

[30] J. N. TSITSIKLIS, *Problems in decentralized decision making and computation*, PhD thesis, Massachusetts Institute of Technology, 1984.

[31] J. N. TSITSIKLIS, D. P. BERTSEKAS, AND M. ATHANS, *Distributed asynchronous deterministic and stochastic gradient optimization algorithms*, IEEE Transactions on Automatic Control, 31 (1986), pp. 803–812.

[32] G. VARATKAR, S. NARAYANAN, N. SHANBHAG, AND D. JONES, *Sensor network-on-chip*, in International Symposium on Systems-on-Chip, 2007, pp. 1–4.

[33] L. XIAO, S. BOYD, AND S.-J. KIM, *Distributed average consensus with least mean square deviation*, Journal of Parallel and Distributed Computing, 67 (2007), pp. 33–46.

[34] N. YE, ed., *Data mining handbook*, Lawerence Erlbaum Associates, 2002, ch. Distributed data mining: Algorithms, systems, and applications. Authored by B. Park and H. Kargupta.